AD_____

AWARD NUMBER: W81XWH-05-1-0562

TITLE: Molecular and Clinical Predictors of Aggressive Prostate Cancer

PRINCIPAL INVESTIGATOR: Lorelei A. Mucci

CONTRACTING ORGANIZATION: Brigham and Women's Hospital, Inc.
Boston, MA 02115

REPORT DATE: September 2008

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1 September 2008	Final	1 Sep 2005 – 31 Aug 2008

4. TITLE AND SUBTITLE

Molecular and Clinical Predictors of Aggressive Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-05-1-0562

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Lorelei A. Mucci

E-Mail: lmucci@hsph.harvard.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Brigham and Women's Hospital, Inc.
Boston, MA 02115

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
While prostate cancer is an important cause of cancer mortality, most men diagnosed with early prostate cancer experience an indolent course. We evaluated molecular and clinical predictors to distinguish lethal and indolent prostate cancer. In a related project, we tested the predictive value of a previously identified multigene tumor signature, a 12-gene model. Risk classification based on the 12-gene model predicted development of lethal disease 20 years hence. The best discrimination came from combining data from the 12-gene markers and clinical data, which perfectly classified the lowest risk stratum where no one died of cancer and provided greater discriminatory ability (AUC 0.78) than the clinical model alone (AUC 0.71), p=0.04. We tested the top 5-genes from the model, and additional tumor markers, within a prostatectomy cohort of 950 men from the Physicians' Health Study and Health Professionals Follow-up Study. We constructed high-density tumor tissue microarrays, and undertook immunohistochemistry to characterize protein expression. We abstracted clinical data from medical records and pathology reports. We undertook statistical analyses and model building to test the discriminatory ability of the gene markers and clinical data, with the ultimate goal to provide prognostication of lethal and indolent prostate cancer.

15. SUBJECT TERMS
Biomarker, Prostate Cancer Survival, Statistical Models, Aggressive Prostrate

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU		19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

Upon diagnosis with localized prostate cancer, patients and clinicians are faced with the decision of whether to treat or to defer treatment. On one hand, prostate cancer is a leading cause of cancer death among men in westernized countries (1), and deaths occur even 20 years after diagnosis (2). On the other, treatment has adverse effects (3) and is often unneeded, as most men do not die from their cancer, and many harbor tumors which are indolent even in the absence of therapy (2, 4, 5). Thus, there is a clear need for tools to distinguish potentially lethal from indolent disease at diagnosis to guide treatment decisions. In the current project, we seek to test molecular and clinical predictors at diagnosis to discriminate lethal and indolent prostate cancer within two large US cohorts of prostate cancer cases from the Physicians' Health Study (PHS) and the Health Professionals' Follow-up Study (HPFS). In Aim 1, we will develop the best set of clinical markers to predict aggressive prostate cancer, as defined by development of distant metastases or prostate cancer death. In Aim 2, we will develop a composite biomarker to predict aggressive prostate cancer, and quantify the extent to which these markers, in combination with clinical parameters, can discriminate aggressive from indolent disease.

BODY

Aim I. To determine clinical predictors of aggressive prostate cancer.

We completed abstraction of clinical data from medical records and pathology reports for the 6,040 prostate cancer cases in the Physicians' Health Study (PHS) and Health Professionals' Follow-up Study (HPFS. We have retrieved information on tumor grade, stage, extent, PSA levels at diagnosis, clinical presentation and treatments.

A clinical database in SAS was constructed from these data, and quality control assessment of this database is finalized. The clinical database from the prostate cancer cases has been merged with additional data files, including death files and questionnaire data, from the PHS and HPFS cohorts.

For approximately 1,100 cases in the PHS and HPFS cohorts, we have collected information on Gleason grade through a standardized Gleason scoring by the pathologist on our project, and have incorporated these data into a SAS dataset on the UNIX system.

We have completed a statistical analysis to develop a model of clinical parameters that predicts prostate cancer death. Using Cox proportional hazard models, we have found Gleason grade at diagnosis to be one of the strongest predictors of lethal and indolent prostate cancer, with relative risks of 3.8 ($p<0.001$) and 7.1 ($p<0.001$) for Gleason 7 and Gleason 8-10 tumors compared to Gleason 2-6. Among men with Gleason 7 based on the original review, there was no prognostic difference between Gleason 3+4 versus 4+3 tumors, an observation noted by Andren et al(6). We also oberved a significant association between older age at diagnosis ($p<0.001$) and higher psa levels at diagnosis ($p<0.001$) with development of lethal prostate cancer.

However, we see big differences comparing the original and re-review. There was generally an upgrading in Gleason comparing the original to the re-review score, and almost 90% of cancers originally scored as 2-5 were rescored as 6 or 7 tumors. The discrimination of models of standardized scores from prostatectomy (C-statistic: 0.86) and biopsy (C-statistic=0.85) were improved compared to models of original scores (prostatectomy C-statistic: 0.82; biopsy C-statistic: 0.72). Moreover, 4+3 cancers were associated with a three-fold increase in lethal prostate cancer compared to 3+4 cancers (95% CI: 1.1, 8.6). **(Table 1**) Ignoring the predominance of Gleason pattern 4 in GS 7 cancers may conceal important prognostic information. A standardized review of GS can improve prediction of PCa survival. These findings were presented at the 2008 AACR Meeting in San Diego 2008, and the manuscript is accepted for publication in *Journal of Clinical Oncology* (Stark JR et al, Gleason score and Lethal Prostate Cancer: Does 3+4 = 4+3?, J Clin Oncol 2009 In press **Appendix 1**).

Table 1. Comparison of re-review and original Gleason score to predict lethal prostate cancer, PHS and HPFS Cohort 1982-2008

	Re-reviewed Gleason Score				Original Gleason Score			
Total Gleason	PrCa Deaths	Person Yrs	Mortality Rate (per 1,000 p-yrs)	Hazard Ratio (95% CI)[a]	PCa Deaths	Person Yrs	Mortality Rate (per 1,000 p-yrs)	Hazard Ratio (95% CI)[a]
2-5	0 64.6		0	No deaths	1	2178.8	0.5	0.1 (0.0, 0.5)
6	0 2216.0		0	No deaths	3	2331.5	1.3	0.2 (0.1, 0.6)
3+4	6 2864.9		2.1	1.0 (Ref)	12 1701.8		7.1	1.0 (Ref)
4+3	9	1419.1	6.3	3.1 (1.1, 8.6)	9	542.5	16.6	2.4 (1.0, 5.6)
8	7	482.3	14.5	7.4 (2.5, 22.0)	4	435.3	9.2	1.3 (0.4, 4.0)
9-10	15	383.7	39.1	19.1 (7.4, 49.2)	8	240.7	33.2	4.9 (2.0, 11.9)
Total	37 7430.6		5.0	-	37	7430.6	5.0	-

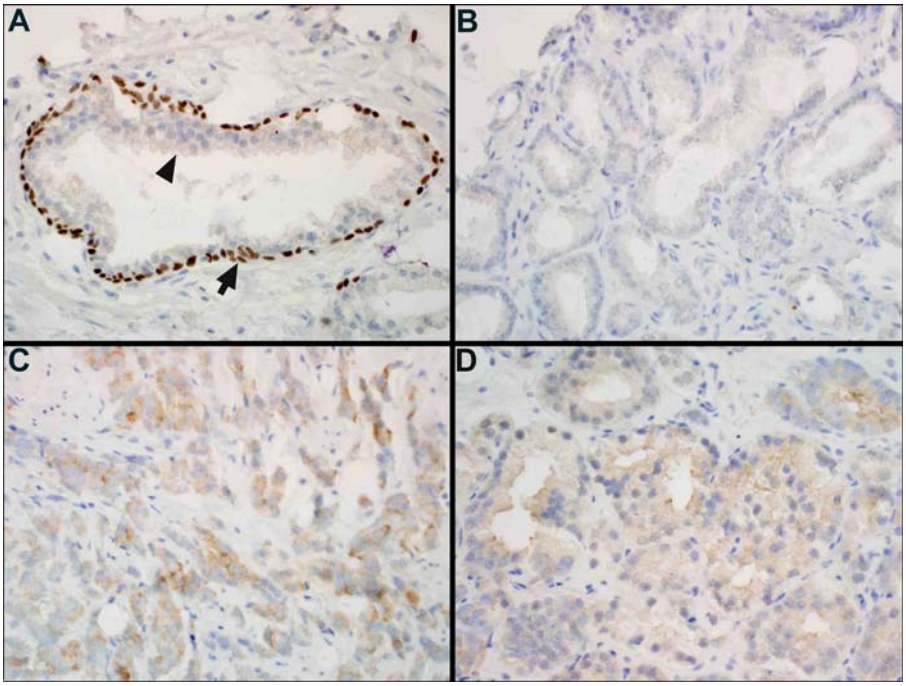
Aim II. To develop a composite set of molecular predictors of aggressive prostate cancer by a multiplex approach.

We have collected archival tumor tissue materials for 1,500 HPFS cases and PHS cases, and a tissue tracking system has been developed for the PHS and HPFS tumor repository. The prostate cancer cases in the PHS and HPFS tumor repository have undergone pathologic review to obtain standardized review of Gleason grade, tumor extent, tumor type, and other histological features. We have completed construction of 10 tumor tissue microarrays from prostatectomy specimens and the tissue biomarkers have been completed on these samples.

The biomarkers, p63, MTA1, Jagged1, AMACR, and ciap1 have been stained by immunohistochemistry and evaluated for intensity and protein expression using the Chromavision Imaging system. In 2007, we began using the Ariol Imaging system for quantitative immunohistochemistry, and have completed staining on TMAs for KI67 and apoptosis (TUNEL) using this system.

P 63. The predominant location of tumor p63 staining in our samples occurred in the cytoplasm (**Figure 1**), which is uncommon and reflects a departure from the strong nuclear staining usually observed in non-neoplastic basal cells. Increasing expression of cytoplasmic p63 was associated with prostate cancer mortality (n=19 deaths); the hazard ratios from tertiles 1-3 were: 1.0 [reference], 4.0 (95% CI: 0.9, 18.9) and 5.9 (95% CI: 1.3, 27.5), p for trend = .03). The positive trend remained significant (p=.047) after multivariable adjustment for age, year of diagnosis and Gleason score. Higher tertiles of cytoplasmic p63 were also associated with reduced levels of apoptosis (p for trend = 0.0408) and increased cellular proliferation (p for trend = 0.0026).
We found aberrant expression of p63 in the cytoplasm to be associated with increased prostate cancer-specific mortality up to 20 years after diagnosis. The mislocalized expression was associated with reduced apoptosis and higher proliferative activity, and may suggest an oncogenic role in prostate cancer progression and survival. These findings were published in *Cancer Epidemiology, Biomarkers and Prevention* (Dhillon P et al, Aberrant cytoplasmic expression of p63 and prostate cancer mortality. CEBP 2008).

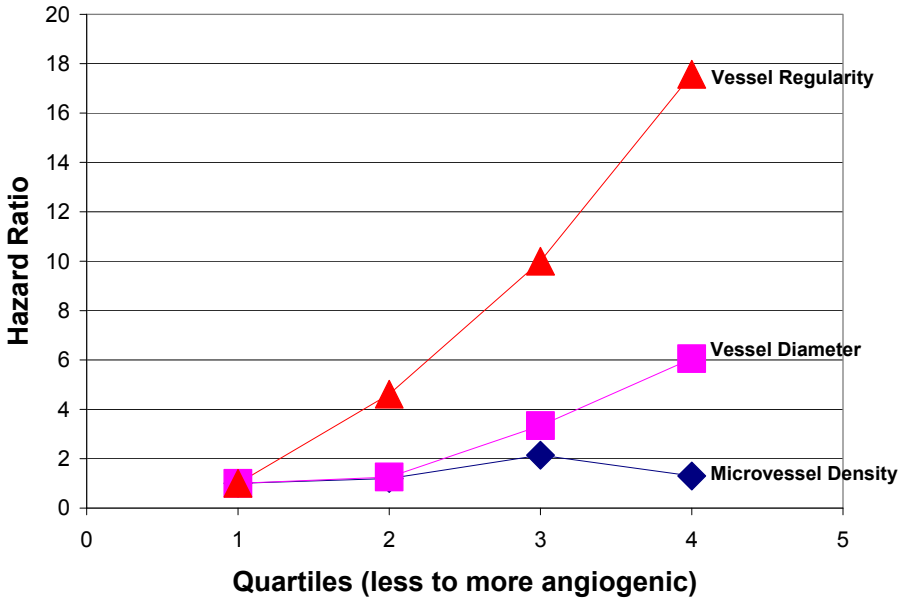
Figure 1. (A) Normal prostate gland showing p63 staining of basal cells (arrow) and lack of staining of luminal cells (arrowhead) (x40). (B) Prostate adenocarcinoma showing no cytoplasmic or nuclear staining for p63 (x40). (C & D) Prostate adenocarcinoma showing cytoplasmic staining for p63 (x40, x100).



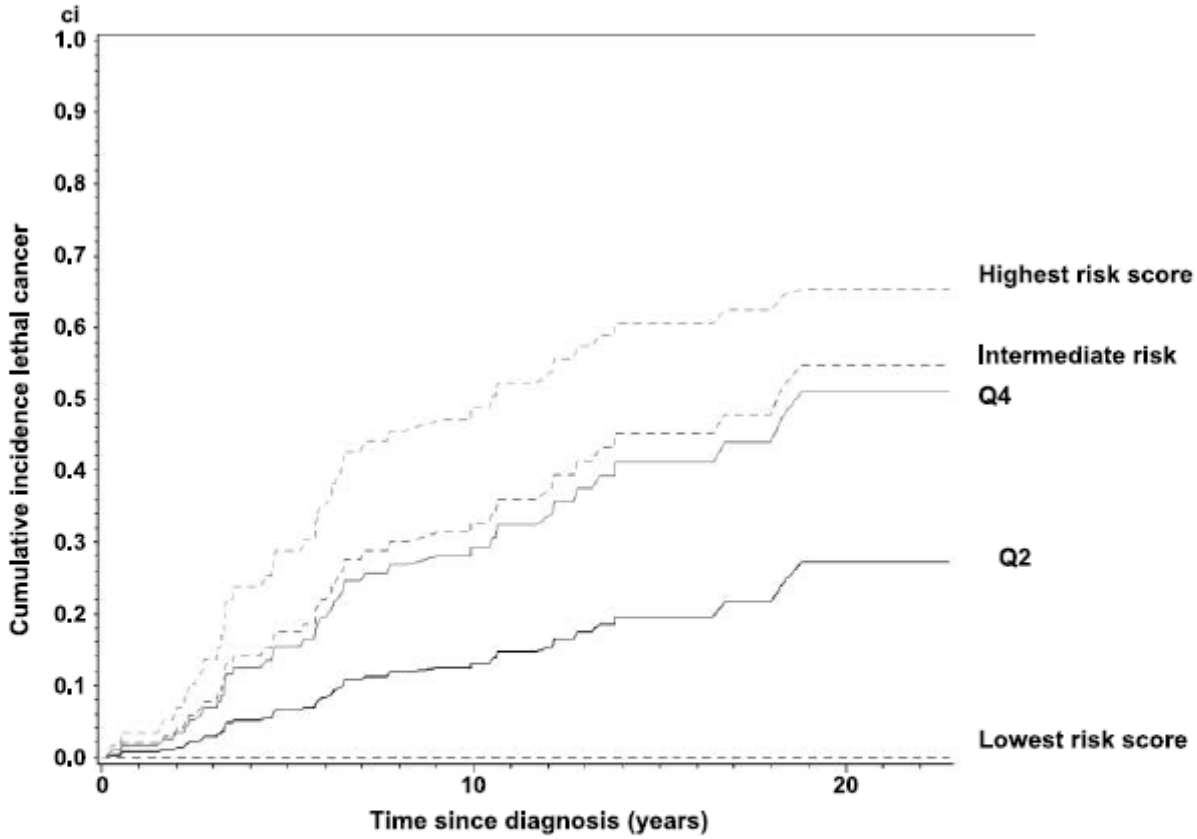
cIAP1. The cellular inhibitor of apoptosis protein 1 is part of a family of proteins that regulate apoptosis. Lower cIAP1 staining was associated with reduced tumor apoptosis (p-for-trend 0.003), with more advanced pathological tumor stage (p-for-trend 0.0004), and higher Gleason score (p-for-trend 0.02). Moreover, lower CIAP expression was associated with increased risk of developing lethal prostate cancer during an average 10 years of follow-up, independent of clinical factors (HR = 1.7; 95% CI 1.0-2.8). The association was strongest among tumors with Gleason grade ≤7 (HR=4.7; 95% CI 1.4-15.6). This study adds to the understanding of cIAP1 regulation in prostate cancer progression, through apoptotic dependent and independent mechanisms. A manuscript summarizing these findings is under review by a cancer journal.

MTA1. The metastasis-associated gene 1 (MTA1) is part of a family of MTA proteins that function as coregulators of nuclear receptors, modulating transcription of target genes. MTA1 is overexpressed in several human cancers, including prostate cancer, and is highly upregulated in metastatic versus localized prostate cancer tissue. These characteristics support MTA1's role in cancer metastasis. We explored the relation between protein expression of MTA1 in tumor tissue and clinical characteristics and risk of lethal prostate cancer. In addition, we correlated MTA1 expression to morphologic assessment of tumor angiogenesis and cellular proliferation. Linear regression was used to evaluate the association between MTA1 quartiles and angiogenesis markers. MTA1 protein expression showed a moderate positive correlation with the nuclear proliferation marker, ki67 (r=0.33, p-value=<0.0001). After adjustment for age at diagnosis and Gleason score, MTA1 expression showed a weak inverse correlation with vessel size (r=-0.14, p=0.04) and vessel diameter (r=-0.13, p=0.06). Adjusting for Gleason score and age at diagnosis, there was a statistically significant 3-fold increased risk of lethal prostate cancer in the highest quartile of MTA1 expression compared to the lowest quartile.

Angiogenesis. Tumor growth requires the development of independent vascular networks that are often primitive in morphology and function. We examined whether microvessel morphology contributes to the considerable biologic heterogeneity of prostate cancer. We evaluated microvessel morphology as a predictor of prostate cancer mortality among 572 men in the Health Professionals Follow-up Study diagnosed with cancer during 1986-2000. We immunostained prostatectomy tumor block sections for endothelial marker CD34, and assessed microvessel density, vessel size (area and diameter), and irregularity of vessel lumen using image analysis. Proportional hazard models were used to assess microvessel density and morphology in relation to lethal prostate cancer. Poorly differentiated tumors exhibited greater microvessel density, greater irregularity of the vessel lumen, and smaller vessels. During 20 years of follow-up, 44 men developed bone metastases or died of cancer. Men with tumors exhibiting the smallest vessel diameter, based on quartiles, were 6.0 (95% CI, 1.8-20.0) times more likely to develop lethal prostate cancer **Figure 2**. Men with the most irregularly shaped vessels were 17.1 (95% CI 2.3-128) times more likely to develop lethal disease. Adjusting for Gleason grade and PSA levels did not qualitatively change the results. Microvessel density was not linked to cancer-specific mortality after adjusting for clinical factors. Aggressive tumors form vessels that are primitive in morphology and function with consequences for metastases. Vascular size and irregularity reflect prostate cancer's angiogenic potential, and may serve as biomarkers to predict prostate cancer mortality several years after diagnosis. These data were included in a manuscript accepted for publication in J Clin Oncology.



Multigene signatures. We undertook a complementary analysis of tumor biomarkers within the Swedish Watchful Waiting Cohort, among men who were diagnosed with prostate cancer and followed initially by active surveillance. We tested in this cohort whether a 12-gene or 9-gene model previously identified through expression array analyses would predict lethal prostate cancer. We found that classifying men based on expression of genes in the 12-gene model provided excellent discrimination of men with a lethal vs. indolent phenotype. **Figure 3** below presents the cumulative incidence of lethal prostate cancer during 20 years of follow-up comparing men with a high to low risk of progression based on the 12-gene model. These findings were published in Cancer Epidemiology, Biomarkers and Prevention. This study provides additional support for the current project. In contrast, the 9-gene model did not predict outcome.



	Follow-up time			
Follow-up	5 years	10 years	15 years	20 years
N at risk	106	69	26	4
CI difference				
Lowest risk	REF*	REF*	REF*	REF*
Highest risk	+28.7% (17.4-40.0)	+48.7% (37.2-60.2)	+60.4% (50.3-70.5)	+65.4% (57.2-73.6)

* The lowest risk group is the referent group

KEY RESEARCH ACCOMPLISHMENTS

- Completed standardized review of cohort for Gleason grading, and finalized construction of all tumor tissue microarrays
- Completed biomarker studies on tumor tissue microarrays
- Created a merged clinical and tissue SAS database, including clinical information, and biomarker data for statistical analyses
- Data presented as poster and oral presentations at the American Association for Cancer Research Meeting in San Diego (2008) and the AACR
- Published 5 journal articles; 1 journal article under review; 2 journal articles in preparation

REPORTABLE OUTCOMES

- Student working on this project (Jennifer Stark) was promoted to post-doctoral fellow
- Dr. Stark received a Career Development Award from the Dana Farber/Harvard Cancer Center Prostate Cancer SPORE based on an extension of this project
- Dr. Mucci received a Young Investigators Award from the Prostate Cancer Foundation based on experience supported by this award
- Dr. Mucci was promoted to Assistant Professor of Medicine and Epidemiology based on this award
- Development of prostate tumor tissue repository for 1,500 participants in the PHS and HPFS cohorts
- Data presented at national cancer meetings (American Association of Cancer Research 2008; AACR Cancer Prevention Meeting 2008)
- Data included in 5 published manuscripts; 3 in preparation/under review

CONCLUSION

We have demonstrated our ability to undertake this large cohort and collect archival tumor specimens from 680 men. We have demonstrated a proven working relationship with the pathology team, as shown by substantial progress on construction of the tissue microarrays, standardized Gleason grading, evaluation of atrophy and inflammation, and successful completion of two biomarkers on the tissue microarrays. Moreover, our preliminary statistical analyses on atrophy-inflammation and prostate cancer-specific mortality, combined with the findings of serologic evidence of *T vaginalis* and prostate cancer mortality provide supportive evidence for the study hypothesis.

During the next year, we will complete the remaining biomarker analyses and undertake statistical analyses to successfully complete the study aims. Ultimately, this project has the potential to provide strong evidence (for or against) in assessing the role of infectious agents and inflammation in prostate pathogenesis. Moreover, the tumor tissue repository we are establishing is a unique resource in which to test future hypothesis. Given the substantial biologic heterogeneity of prostate cancer, the proposed project would ultimately have exciting implications for prevention and potentially treatment of prostate cancer.

REFERENCES

1. Ferlay J, Bray F, Pisani P, Parkin DM. GLOBCAN 2000: Cancer Incidence, Mortality and

Prevalence Worldwide. Lyon: IARCPress, 2001.

2. Steinek G, Helgesen F, Adolfsson J, et al. Quality of life after radical prostatectomy or watchful waiting. New England Journal of Medicine 2002;347:790-796.

3. Holmberg L, Bill-Axelson A, Helgesen F, et al. A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer. New England Journal of Medicine 2002;347:781-789.

4. Nelson WG, De Marzo AM, DeWeese TL, Isaacs WB. The role of inflammation in the pathogenesis of prostate cancer. The Journal of Urology 2004;172:S6-S12.

5. Kuper H, Adami HO, Trichopoulos D. Infections as a major preventable cause of human cancer. Journal of Internal Medicine 2000;248:171-183.

6. Palapattu GS, Sutcliffe S, De Marzo AM, Isaacs WB, Nelson WG. Prostate carcinogenesis and inflammation: emerging insights. Carcinogenesis 2004;26:1170-1181.

7. Irani J, Goujon J-M, Ragni E, et al. High-grade inflammation in prostate cancer as a prognostic factor for biochemical recurrence after radical prostatectomy. Urology 1999;54:467-472.

8. Kuper H, Adami H-O, Trichopoulos D. Infections and a major preventable cause of human cancer. Journal of Internal Medicine 2000;248:171-183.

9. Pisani P, Parkin DM, Munoz N, Ferlay J. Cancer and infections: estimates of the attributable fraction in 1990. Cancer Epidemiology Biomarkers and Prevention 1997;6:387-400.

10. Signorello LB, Adami HO. Prostate Cancer. In: Adami HO, Hunter DJ, Trichopoulos D, eds. Textbook of Cancer Epidemiology. New York: Oxford University Press, 2002:385.

11. Sutcliffe S, Giovannucci E, Alderete JF, et al. Plasma antibodies against Trichomonas vaginalis and subsequent risk of prostate cancer. Cancer Epidemiology Biomarkers and Prevention 2006;15: 939-45.

12. Urisman A, Molinaro RJ, Fischer N, et al. Identification of a novel gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. PLOS Pathogens 2006;2:e25.

APPENDICES

1. Manuscript published in J Clinical Oncology on Gleason grading

In press: Journal of Clinical Oncology 2009

Gleason Score and Lethal Prostate Cancer: Does 3+4 = 4+3?

Jennifer R. Stark, ScD[#]; Sven Perner, MD, PhD; Meir J. Stampfer, MD, DrPH; Jennifer A. Sinnott, MS; Stephen Finn, MD, PhD; Anna S. Eisenstein, BS; Jing Ma, MD, PhD; Tobias Kurth, MD, ScD; Massimo Loda, MD; Edward L. Giovannucci, MD, DrPH; Mark A. Rubin, MD; Lorelei A. Mucci, ScD

Author Affiliations: Departments of Epidemiology (Drs. Stark, Stampfer, Kurth, Giovannucci, and Mucci), Biostatistics (Ms. Sinnott), and Nutrition (Drs. Stampfer and Giovannucci), Harvard School of Public Health, Boston, Massachusetts; Channing Laboratory (Drs. Stark, Stampfer, Giovannucci, Ma, and Mucci, and Ms. Eisenstein) and Division of Preventive Medicine (Dr. Kurth), Department of Medicine, Brigham and Women's Hospital, Boston; Department of Pathology and Laboratory Medicine, Weill Medical College, Cornell University, New York, New York (Drs. Perner and Rubin); Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston (Drs. Finn and Loda)

Funding/Support: This work was supported by grants from the National Cancer Institute (5R01CA058684-13 and 5R01CA042182-20) and Department of Defense (W81X WH-05-1-0562). The Physicians Health Study is supported by grants CA34944, CA40360, and CA097193 from the National Cancer Institute and grants HL-26490 and HL-34595 from the National Heart, Lung, and Blood Institute. Dr. Stark is supported by the National Research Service Award Training Program in Cancer Epidemiology (T32 CA009001-32) and a Dana-Farber/Harvard Cancer Center SPORE Career Development Award.

Running title: Gleason score and lethal prostate cancer

Previous Presentation: An abstract based on this manuscript was presented at the American Association for Cancer Research Annual Meeting; April 12-16, 2008, San Diego, California.

Disclaimers: None

Abstract

Purpose: Gleason grading is an important predictor of prostate cancer (PCa) outcomes. Studies using surrogate PCa endpoints suggest outcomes for Gleason score (GS) 7 cancers vary according to the predominance of pattern 4. These studies have influenced clinical practice, but it is unclear if rates of PCa mortality differ for 3+4 and 4+3 tumors. Using PCa mortality as the primary endpoint, we compared outcomes in Gleason 3+4 and 4+3 cancers, and the predictive ability of GS from a standardized review versus original scoring.

Patients and Methods: Three study pathologists conducted a blinded standardized review of 693 prostatectomy and 119 biopsy specimens to assign primary and secondary Gleason patterns. Tumor specimens were from PCa cases diagnosed between 1984 -2004 from the Physicians' Health Study and Health Professionals Follow-up Study. Lethal PCa (53 cases) was defined as development of bony metastases or PCa death. Hazard ratios (HR) were estimated according to original GS and standardized GS. We compared the discrimination of standardized and original grading with C-statistics from models of 10-year survival.

*Result*s: For prostatectomy specimens, 4+3 cancers were associated with a three-fold increase in lethal prostate cancer compared to 3+4 cancers (95% CI: 1.1, 8.6). The discrimination of models of standardized scores from prostatectomy (C-statistic: 0.86) and biopsy (C-statistic=0.85) were improved compared to models of original scores (prostatectomy C-statistic: 0.82; biopsy C-statistic: 0.72).

Conclusion: Ignoring the predominance of Gleason pattern 4 in GS 7 cancers may conceal important prognostic information. A standardized review of GS can improve prediction of PCa survival.

Introduction

Gleason grading[1] is a strong predictor of survival among men with prostate cancer (PCa).[2] The Gleason system, introduced in 1974,[3] is an architectural grading system that ranges from 1 (well differentiated) to 5 (poorly differentiated). The Gleason score (GS) is the sum of the primary and secondary patterns with a range of 2 to 10. It has long been appreciated that patients with GS $\geq$ 7 are at greater risk for extraprostatic extension and biochemical recurrence.[4] However, more recent evidence suggests that the application of GS has changed, leading to changes in the distribution of scores over time.[5-7] Further, studies using surrogate endpoints have shown that the prognosis of GS 7 cancers varies considerably.[8-10]

Analyses of GS assignment before and after the introduction of PSA screening,[7,11,12] and studies that have involved blind re-reviews of original specimens,[5-7,13] have noted increases in GS with more contemporary readings. The change in GS over time appears to be largely due to a systematic shift in Gleason grading, which is attributed primarily to two factors.[14] First, it is widely accepted that Gleason grade from biopsy is frequently upgraded at prostatectomy, resulting in a reluctance to assign a low GS at diagnosis.[15] Second, previously undescribed benign lesions (i.e., atypical adenomatous hyperplasia) were historically mistaken for Gleason 1+1 tumors, but contemporary scoring more often correctly classifies those lesions as benign.[16] Because systematic upgrading in Gleason results in improved survival for all Gleason categories, the observed Gleason shift has been described as the “Will Rogers” phenomenon.[5,7]

Historically, PCa risk prediction models[4,17] and observational studies that have adjusted for GS utilized only the overall GS. Because numerous studies suggest that Gleason 3+4 tumors (i.e., those where pattern 3 is most prevalent but some amount of pattern 4 is also observed) have a better prognosis than Gleason 4+3 tumors (i.e., those where pattern 4 is more prevalent than pattern 3), contemporary clinical risk prediction now incorporates primary and secondary Gleason pattern.[18] However, studies that have specifically assessed differences in Gleason 3+4 and 4+3 cancers have relied on associations of Gleason pattern with other prognostic factors and biochemical progression[8,10,12,19-21] or, less frequently, development of metastases.[8,21] While biochemical recurrence is a widely used outcome for studies relating to PCa risk prediction and treatment efficacy, it is important to note that its definition varies,[22,23] and it is an imperfect surrogate for PCa mortality.[24,25] We undertook a standardized review of radical prostatectomy and needle biopsy tumor tissue samples from the Physicians’ Health Study (PHS) and the Health Professionals Follow-Up Study (HPFS) PCa

cohorts to assess the predictive ability of a contemporary review compared to original Gleason scoring, as well as explore potential differences in GS 7 subtypes, with PCa mortality as the primary endpoint.

Patients and Methods

Patient Cohort. The PHS [26-28] was initiated as a randomized, double-blind, placebo-controlled trial for the primary prevention of cardiovascular disease and cancer. The study included 29,021 healthy U.S. male physicians age 40 to 84 years at baseline. The majority of participants were Caucasian (94%). Participants are followed through annual questionnaires to collect data on diet, health and lifestyle behaviors, and medical history, and biannually to ascertain compliance and health endpoints, including PCa.

The HPFS was initiated in 1986, when 51,529 male health professionals, ages 40 to 75, completed a mailed questionnaire on demographic characteristics, risk factors and preventive behaviors, and diet and use of supplements. The cohort is predominantly Caucasian (>91%). Through biennial follow-up mailed questionnaires, we update exposure information and medical events, including PCa.

In both cohorts, participants are asked to report new diagnoses of cancer on follow-up questionnaires. Subsequently, hospital records and pathology reports are requested and reviewed by study investigators. Through systematic medical record review of PCa cases, we abstract clinical and pathological data. When available, the original Gleason score, major and minor pattern are recorded. Stage is recorded according to the TNM staging system or a modified Whitmore-Jewett classification scheme. We also follow PCa cases through questionnaires to collect information on their PCa clinical course, including PSA levels and development of metastases. Deaths are ascertained through mailings, telephone calls and periodic searches of the National Death Index, and cause of death is assigned after review of death certificates, information from the family, and medical records. Follow-up for mortality is >99% complete in the PHS and >98% complete in the HPFS.

For PHS cases, original specimens (blocks and H&E slides) from needle biopsy or prostatectomy are requested from the diagnosing institution of every PCa case. For HPFS cases, only prostatectomy specimens are requested. Of the 1,195 tissue blocks obtained, we excluded two cases determined not to have cancer upon standardized review, as well as two cases found to have transitional cell bladder cancer. An additional 379 cases were excluded because an original Gleason score from the same specimen type as the obtained tissue was not available. A total of 812 men with PCa were included in the study. Blinded to the original

pathology reports and clinical data, we undertook a standardized review (MAR, SP, SF) of original H&E slides from the referring hospitals and assigned a primary and secondary Gleason grade. The pathologists reviewed the slides independently, and then Gleason scores were compared. For any cases with discrepant Gleason data between pathologists, slides were re-reviewed until a consensus was reached.

Statistical Analysis. All analyses were conducted separately for prostatectomy and needle biopsy specimens. To assess differences in original and standardized GS, we compared distributions among the men using a Wilcoxon signed-rank test. Trends in Gleason during three four-year time periods (1985-1988, 1994-1997, and 2001-2004) were explored by plotting the original vs. standardized scores and the best-fitting regression line through the points. To compare the ability of GS from original report and standardized review to predict PCa survival, the outcome event was defined as lethal PCa. Event dates were the date of diagnosis of bony metastases when such data were available, or the date of PCa death, otherwise. Cases who did not die of PCa and who did not develop metastases to bone were censored at time of death from other causes or the end of study follow-up (March 1, 2008). Follow-up time was calculated from the date of PCa diagnosis to the event date, or time of censoring.

Crude mortality rates were calculated within six strata defined by the original reports and standardized GS (2-5, 6, 3+4, 4+3, 8, and 9-10). Time-to-event analyses to predict lethal PCa were also conducted separately for the original and the standardized Gleason data. Cox proportional hazards models that controlled for age at diagnosis were used to estimate hazard ratios (HR) and 95% confidence intervals (95% CI) by including in the model indicator variables for each stratum of Gleason.

The discrimination of models that included the GS based on the standardized review and historical grading was compared using 10-year survival as an endpoint. In this analysis, we contrasted two groups: men who lived at least 10 years following diagnosis without known metastases, and men who developed metastases or died of PCa within ten years after diagnosis. To obtain C-statistics for both original and standardized scores, a 10-level ordinal variable for GS 2-10 (with separate codes for Gleason 3+4 and 4+3) was included in a logistic regression model. We fit crude models and models that adjusted for age, stage, and PSA at diagnosis. We obtained 95% confidence intervals (CI) by repeating the analysis on 1,000 bootstrap

samples.[29] All analyses were performed using SAS version 9.1.3. The research protocol was approved by the institutional review board at the Harvard School of Public Health and Partners Healthcare.

Results

Selected characteristics of the study population are included in Table 1. The mean age at diagnosis was 65.5 years among the 693 cases with prostatectomy specimens and 71.2 years among the 119 cases with biopsy specimens. Most cases were diagnosed in the PSA screening era (85% of prostatectomy and 93% of biopsy specimens).

We observed a dramatic and statistically significant shift in Gleason grading upon standardized review for both prostatectomy ($p<0.0001$) and biopsy specimens ($p<0.0001$). The shift in GS assignment of prostatectomy specimens is illustrated in Table 2. Among the prostatectomy cases, 171 cases were originally assigned a GS of 2-5, but upon standardized review, only 6 cases had that assignment (all of whom were GS 5). At the upper end of the scale, 28 prostatectomy cases were originally assigned Gleason 9 or 10, compared to 45 upon standardized review. For biopsy specimens, GS assignment of 2-5 decreased from 20 to 1 and GS of 9-10 decreased from 7 to 3.

Figure 1 illustrates the trends in Gleason grading of prostatectomy specimens during the study period. Our pathologists tended to give scores that were higher than those originally assigned from 1985-1988. Compared to the scores assigned in 1994-1997, the standardized GS at the upper end of the scale were generally upgraded from their original score. This trend appeared to persist through 2001-2004. When we divided the data into three time periods (Pre-1994, 1994-2000, and 2001-2004), the weighted Kappa statistics for concordance between original and standardized scores increased in each subsequent time period (from 23%, to 33%, to 44%), while the percent of cases with original GS that were lower than the standardized scores simultaneously decreased (from 73%, to 55%, to 39%).

A total of 53 lethal PCa events, which included the development of bony metastases (N=11) and PCa-specific deaths (N=42), occurred during the study period. Of the 241 total cases (biopsy and prostatectomy) assigned GS 2-5 or 6 upon standardized review, none developed lethal PCa. Among the 693 prostatectomy patients, 37 developed lethal PCa, with post-diagnostic survival time ranging from 0.1-21.1 years (median

10.4). As shown in Table 3, crude cancer mortality rates increased in each consecutive stratum of standardized GS for the prostatectomy specimens.

Since no lethal events occurred among prostatectomy cases scored as Gleason 2-6 in the standardized review, we designated 3+4 as the reference group when estimating HRs for lethal PCa. After adjusting for age at diagnosis, rates of lethal PCa for prostatectomy specimens assigned as GS 4+3, 8, and 9-10 upon standardized review were significantly higher than Gleason 3+4 (Table 3), and rates increased with each level of Gleason. Among those with an prostatectomy specimen, cases with a standardized GS of 4+3 were 3.1 times more likely to develop lethal PCa than cases with 3+4 (94% CI: 1.1, 8.6), while cases with a standardized GS of 9-10 were more than 19 times more likely to develop lethal PCa compared to cases with GS of 3+4 (HR: 19.1; 95% CI: 7.4, 49.2). Using the original GS from prostatectomy, an elevated rate of lethal PCa was also found when we compared GS 4+3 vs. 3+4 (HR: 2.4; 95% CI: 1.0, 5.6). Cases with prostatectomy specimens assigned GS 9-10 in the original report were only 4.9 times as likely to develop lethal PCa compared to cases with GS 3+4 tumors assigned by the original pathologist (95% CI: 2.0, 11.9) (Table 3). After additional adjustment for pathological stage and log(PSA) at diagnosis among the subset of 502 cases for whom data was available, the point estimates for the standardized and original GS among prostatectomy cases were similar, albeit with wider confidence intervals. The HRs (95% CI) for standardized GS of 4+3, 8, and 9-10 cancers compared to 3+4 were 2.6 (0.4, 16.0), 6.2 (0.9, 44.7), and 25.9 (4.7, 145.2), respectively. For the original GS, there were no deaths in the GS 2-5 category among the subset of men with stage and PSA data. HRs (95% CI) comparing each remaining stratum of original GS to Gleason 3+4 were 0.2 (<0.1, 1.8) for Gleason 6, 3.2 (0.9, 11.1) for Gleason 4+3, 0.5 (0.1, 4.3) for Gleason 8, and 3.3 (0.8, 13.7) for Gleason 9-10.

For the 119 patients with biopsy specimens, 16 developed lethal PCa and post-diagnostic survival ranged from 0.1-16.3 years (median 8.3). Crude cancer mortality rates in each stratum of standardized GS of the biopsy specimens were as follows: GS 2-5: No deaths; GS 6: No deaths; GS 3+4: 11.0/1,000 p-years; GS 4+3: 18.7/1,000 p-years; GS 8: 40.6/1,000 p-years; GS 9-10: 98.8/1,000 p-years). According to the original GS of the biopsy specimens, crude cancer mortality rates per 1,000 p-years were 14.0 for GS 2-5, 8.5 for GS 6, 10.8 for GS 3+4, 45.2 for GS 4+3, 26.6 for GS 8, and 15.9 for GS 9-10.

In predicting 10-year survival, we observed a marked improvement in discrimination of models that utilized strata of the standardized scores compared to original scores when modeled alone or with age. For

men with prostatectomy specimens, this analysis included the 380 men who lived at least 10 years following diagnosis without known metastases and the 30 who experienced development of distant metastases or death from PCa within 10 years of diagnosis. When we modeled GS alone, the original score from prostatectomy was associated with a C-statistic of 0.82 (95% CI: 0.77, 0.88), whereas the C-statistic was 0.86 (95% CI: 0.82, 0.91) in the model using the standardized scores (Figure 2). When we modeled GS, age, pathologic stage and log(PSA) at diagnosis, the C-statistics were 0.89 (95% CI: 0.82, 0.96) for the original GS and 0.90 (95% CI: 0.83, 0.97) for the standardized GS. For men with biopsy specimens, the comparison of model discrimination included the 37 men who lived at least 10 years following diagnosis without known metastases and the 12 who developed lethal PCa within 10 years of diagnosis. C-statistics for models of biopsy GS alone were 0.72 (95% CI: 0.58, 0.86) for the original scoring and 0.85 (0.76, 0.94) for the standardized review.

Discussion

As noted in previous studies,[5,7,13] we observed a striking upgrading of GS in a standardized review compared to the original reading. No cases were assigned a GS of 2 upon standardized review and less than 1% received GS 3-5. We observed a 61% increase in the number prostatectomy specimens assigned a GS of 9-10 in the standardized review. More importantly, we found that systematic changes in Gleason grading have improved the ability of GS to predict PCa mortality. Among men with GS <7 assigned in a standardized review, no lethal PCa developed in more than 2,600 person-years of follow-up. Moreover, our analysis revealed that Gleason 4+3 cancers assigned to prostatectomy specimens have three times of the rate of PCa mortality compared to Gleason 3+4 tumors. Considering that capturing the amount of Gleason pattern 4 in tumors revealed remarkable differences in outcome, it is possible that the utilizing the percentage of Gleason 4 may provide further prognostic information, as suggested by a study of biochemical progression [30].

One previous study found that standardized Gleason scores of needle biopsy and TURP specimens were significantly better at discriminating indolent from aggressive disease than the original GS assigned in 1991-1996.[13] Our study confirms these findings in a population where the majority of GS data came from prostatectomy. Interestingly, including data on stage and PSA at diagnosis to our models of PCa mortality produced similar C-statistics for both original and standardized GS (0.89 and 0.90, respectively). These results suggest that stage and PSA levels are more important predictors of PCa mortality when original Gleason data

are used, as would often be the case in epidemiologic studies. As PSA screening continues to increase the number of cases diagnosed with localized cancer and reduces variability in stage at diagnosis,[31] obtaining the most precise assessment of tumor grade will become more critical for population-based studies. Thus, both the primary and secondary Gleason patterns should be considered essential components of prostate cancer data collection for prognostic and research purposes. Given that standard PCa nomograms are most often based on a particular clinical series graded by a single pathologist or one team of pathologists,[18] it is reassuring that GS assigned in a variety of settings do as well as standardized scores in discriminating lethal from indolent cancers when other clinical covariates are considered.

Our findings underscore the difficulty in identifying PCa patients who should be treated with prostatectomy. All men with standardized GS 6 tumors at prostatectomy survived, but many of these men likely would have survived without intervention. By contrast, one third of men with standardized GS 9-10 tumors at prostatectomy developed lethal PCa despite surgery, most likely due to micrometastases at diagnosis. Our biopsy data suggest that, if GS at diagnosis is to be used for guiding treatment decisions, a standardized review is warranted: all 35 men with standardized GS 2-6 at biopsy survived, while 2 of 3 with standardized GS 9-10 at biopsy developed lethal prostate cancer; however, there were no clear trends in mortality rates according to the biopsy GS assigned by diagnosing institutions – a concerning finding given that the original biopsy GS may have factored prominently into disease management.

The present study has several strengths. A large sample of cases with 20 years of follow-up allowed the use PCa mortality as a primary endpoint. In light of data indicating that the risks of dying from PCa and other causes 15 years after biochemical recurrence are virtually equivalent (32% vs. 33%),[32] PSA relapse cannot be viewed as a substitute for more definitive endpoints. The availability of data on primary and secondary Gleason pattern from both the initial review and a standardized review allowed us to utilize mortality data to make comparisons of 3+4 and 4+3 cancers, separately for needle biopsy and prostatectomy specimens. Limitations include missing data on tumor stage and PSA at diagnosis. Nevertheless, our study indicates that contemporary GS from a standardized review can markedly improve prediction of prostate-cancer specific survival compared to the original GS. Further, our study is the first to confirm with prostate-cancer specific mortality data that the predominance of Gleason pattern 4 in GS 7 cancers represents important prognostic information: when it comes to Gleason grading, 3+4 does not equal 4+3.

Acknowledgements

We thank the participants in the Physicians' Health Study and Health Professionals Follow-up Study for their long-standing participation. We are grateful to Julia Fleet, Luba Bondarenko, Al Wing, and Haiyan Zhang for their assistance with data collection and programming.

Table 1. Selected characteristics of study population by source of Gleason data

Characteristics	Prostatectomy (N=693)	Needle Biopsy (N=119)
Age at diagnosis, Mean (±SD) 65.5	(±6.0) 71.2	(±6.4)
Follow-up time, Mean (±SD) 10.7	(±4.0) 8.4	(±3.3)
PSA at diagnosis, Median (Q1-Q3)	6.8 (4.9 – 10.0)	7.2 (5.2 - 12.4)
Pathological Tumor stage, N (%)		
T2	505 (73.9)	-
T3/T4	162 (23.7)	-
N1/M1	16 (2.4)	-
Clinical Tumor stage, N (%)		
T1/T2	-	93 (93.9)
T3/T4/N1/M1	-	6 (6.1)
PSA era, N (%)		
Diagnosed before 1992	103 (14.9)	8 (6.7)
Diagnosed during/after 1992	590 (85.1)	111 (93.3)

Table 2. Comparison of Gleason scores from original source and standardized review among 693 prostatectomy specimens

Original Source		*Standardized Review*									
2		**3**	**4**	**5**	**6**	**3+4**	**4+3**	**8**	**9**	**10**	**Total**
2	0	0	0	0	2	1	0	0	0	0	3
3	0	0	0	0	5	3	0	0	0	0	8
4	0	0	0	0	21	6	1	1	0	0	29
5	0	0	0	3	58	56	9	3	2	0	131
6	0	0	0	2	80	92	37	8	2	0	221
3+4	0	0	0	1	31	73	43	14	9	0	171
4+3	0	0	0	0	1	15	22	12	5	0	55
8	0	0	0	0	1	8	16	10	10	2	47
9	0	0	0	0	1	3	6	2	11	3	26
10	0	0	0	0	0	0	0	1	0	1	2
Total	0	0	0	6	200	257	134	51	39	6	693

Table 3. Prostate cancer mortality rates according to prostatectomy Gleason score

	Standardized Review					*Original Source*				
Total Gleason	PrCa Deaths	Person Yrs	N	Mortality Rate (per 1,000 p-yrs)	Hazard Ratio (95% CI)[a]	PCa Deaths	Person Yrs	N	Mortality Rate (per 1,000 p-yrs)	Hazard Ratio (95% CI)[a]
2-5	0 64.6		6	0	No deaths	1	2178.8	171	0.5	0.1 (0.0, 0.5)
6	0 2216.0		200	0	No deaths	3	2331.5	221	1.3	0.2 (0.1, 0.6)
3+4	6 2864.9		257	2.1	1.0 (Ref)	12 1701.8		171	7.1	1.0 (Ref)
4+3	9	1419.1	134	6.3	3.1 (1.1, 8.6)	9	542.5	55	16.6	2.4 (1.0, 5.6)
8	7	482.3	51	14.5	7.4 (2.5, 22.0)	4	435.3	47	9.2	1.3 (0.4, 4.0)
9-10	15	383.7	45	39.1	19.1 (7.4, 49.2)	8	240.7	28	33.2	4.9 (2.0, 11.9)
Total	37 7430.6		693	5.0	-	37	7430.6	693	5.0	-

[a] From Cox proportional hazards model adjusting for age at diagnosis

Figure 1. Comparison of standardized and original Gleason scores assigned to prostatectomy specimens during three time periods

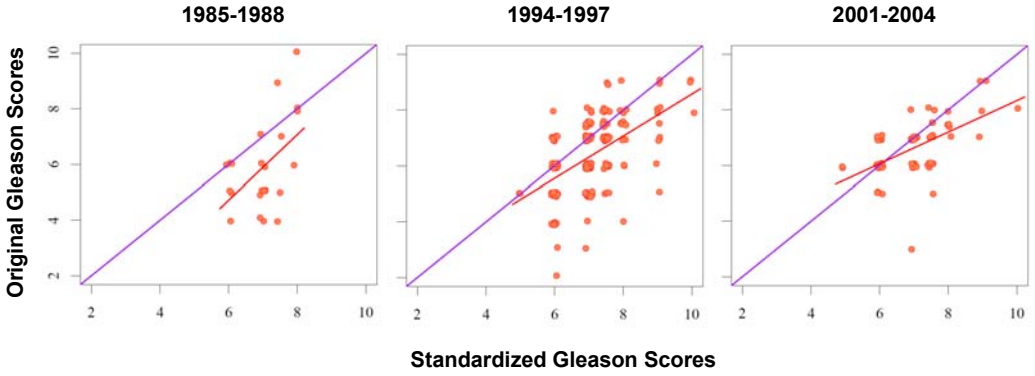


Figure 2. Discrimination of models including Gleason data from original source vs. standardized review of prostatectomy specimens

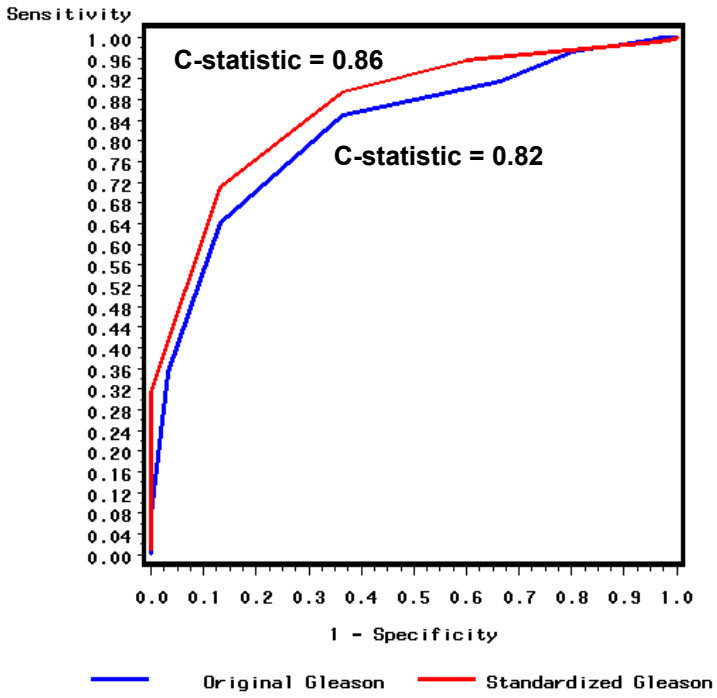


Figure Legends

Figure 1. The Gleason scores and the best fitting line through the points are plotted in red. The purple line represents perfect concordance between original and standardized scores.